

## EL LÉXICO COMO INSTRUMENTO LINGÜÍSTICO DEL SISTEMA UNAM-JURE

Diana CASTAÑEDA PONCE,  
María del Pilar HERNÁNDEZ MARTÍNEZ y  
Jorge Alberto GONZÁLEZ GALVÁN

SUMARIO: I. *Lenguaje, lengua y habla*. 1. *La "lengua" jurídica*. 2. *Los problemas de la "lengua" natural*. II. *Introducción a los sistemas de recuperación de información documental*. III. *El léxico en el sistema UNAM-JURE*. 1. *Definición*. 2. *Objetivo*. 3. *Organización*. A. *Los criterios de organización*. B. *Los problemas lingüísticos*. 4. *Funcionamiento*. A. *Relación de equivalencia*. B. *Relación de extensión*. 5. *Tratamientos especiales*. A. *Por la naturaleza semántica gramatical de las palabras*. B. *Por ser palabras que carecen de parentesco morfológico*. IV. *Conclusiones*.

### I. LENGUAJE, LENGUA Y HABLA

Hablar del léxico dentro del marco de la informática jurídica no es tan fácil como decir: "la lengua es al hombre, como el léxico es a la máquina" (entendido éste como uno de sus instrumentos lingüísticos), ya que el conjugar la ciencia jurídica con los modernos soportes tecnológicos hace que nuestro análisis tome matices que es conveniente detallar, siendo el fenómeno de la comunicación el apoyo de nuestro proyecto.

Describir el proceso de comunicación puede resultar bastante simple si se ve como el diálogo que se entabla entre dos o más sujetos a través de la expresión oral, escrita o mímica; pero trasponer ese mismo fenómeno al hecho de que exista un mecanismo diferente a aquel normalmente empleado, como es la presencia de una máquina, conlleva la necesidad de aclarar ciertos puntos relativos al fenómeno lingüístico, no obstante que la moderna ciencia de la cibernética pretende encontrar los mecanismos propios para la comunicación, representados a través de expresiones matemáticas.

Si bien la materia lingüística se constituye primordialmente por las manifestaciones del lenguaje humano, el estudio del lenguaje —como

capacidad de comunicación de que goza el individuo— presenta dos partes que se corresponden:<sup>1</sup>

- a) La esencial, que tiene por objeto la *lengua*;
- b) La secundaria, que tiene por objeto el *habla*.

La primera es la parte social del lenguaje, independiente del individuo (en cierta medida), pero que es la norma o regla de las demás manifestaciones del lenguaje mismo.

La lengua es un conjunto de convenciones necesarias, adoptadas por el cuerpo social, cuyo ejercicio corresponde a los individuos a través del habla, por lo que la lengua es un sistema de signos que corresponden a ideas distintas, pero que nos permiten identificar y conocer a través del consenso el mundo en que se vive. Dichos signos se pueden comparar con la escritura, las señales de tránsito, el alfabeto de los sordomudos, etcétera, de lo que se infiere que existen diversas clases de signos:

- a) Verbales,
- b) No verbales (v. gr. los gestos) y
- c) Mixtos (v. gr. la ópera).

Es así como se ha concebido la existencia de una “ciencia que estudie la vida de los signos en el seno de la vida social”:<sup>2</sup> la semiología, producto de la fusión de lingüística y psicología social, ya que engloba el sentir del hombre manifestado a través de su lengua, idioma, dialecto o cualesquiera otra forma de comunicación.

Consecuentemente, un análisis semiológico va a consistir en estudiar el propio proceso de comunicación ya aludido, que nos lleva a situarnos ante el acto individual del habla, segundo elemento del lenguaje.

El habla, comúnmente vista o entendida como la facultad de articular palabras, constituye el acto propio del hombre en su individualidad, que históricamente se da con antelación a la lengua, sólo así ésta se establece y evoluciona; sin embargo, la lengua es necesaria para que el habla sea inteligible, de ahí su correspondencia.

La inteligibilidad del habla descansa en que el hombre vive en un mundo de hechos y verdades que capta y asimila con base en fundamentos semióticos, cuya virtud principal es el darle significado y contenido lógico a los signos constituyentes del habla, resaltando a la vista de todos la íntima relación que existe entre lingüística y lógica dentro

<sup>1</sup> Saussure, Ferdinand de, *Curso de lingüística general* (trad. de Mauro Armíño), AKAL, Edit., pp. 35 y 46.

<sup>2</sup> *Idem*, p. 42.

del proceso de comunicación; la primera manifestada mediante la ciencia semiológica, la segunda a través de la semiótica o "doctrina de la naturaleza esencial y de las variedades fundamentales de la semiosis (significación) posible".<sup>3</sup>

Si el *homo sapiens* es también *homo loquens*, se debe precisamente a que es el único ser pensante que por ende puede entender y comprender el sentido de las cosas —su significado—, para luego manejarlo en el proceso de comunicación que requiere para tener vida de un "mensaje", no importando ya los signos que se empleen,<sup>4</sup> sino su esencia.

En este orden de ideas, se puede ya pensar evidentemente que si es el hombre el único ser que cuenta, por obra de la naturaleza, con la facultad de articular palabras, y que consecuentemente él es quien a través de su papel en la sociedad da la pauta para la evolución de la lengua como parte social del lenguaje, también es ya posible el afirmar que existen diversas lenguas, cuya conformación pudiera no ser enteramente natural; pero que en la medida en que el hombre, como ser pensante, juega diversos papeles en el ámbito social, provoca efectivamente que de acuerdo con el grado de civilización alcanzado se desarrollen diversas lenguas especiales, a saber: la terminología del biólogo, del químico, del médico, del músico, del jurista.

Esto explica por qué Eugenio Coseriu se refiere a las lenguas especiales como las "hablas características de grupos sociales o profesionales, como también los lenguajes técnicos".<sup>5</sup>

### I. La "lengua" jurídica

Resulta bastante evidente que hablar en términos de "lengua jurídica" merece una aclaración, puesto que ha sido utilizada de un modo más generalizado (ya como una costumbre) la expresión "lenguaje jurídico", cuando se hace referencia a los signos cuyo contenido semántico es trascendente para la materia jurídica.

Sin embargo, si dicha aplicación ha sido convencionalmente adoptada, es preciso recordar que el lenguaje humano articulado (objeto de la lingüística) consiste, *grosso modo*, en la capacidad que se tiene para expresarse y que su parte social es la lengua o conjunto de signos a los que por consenso se les atribuye un significado.

<sup>3</sup> Cortés, Larrieu, "Semiología, semiótica, teoría de la comunicación", *Revista de Ciencias Sociales*, Chile, 1978, 1er. semestre, p. 253.

<sup>4</sup> *Idem*, p. 255.

<sup>5</sup> Coseriu, Eugenio, *Introducción a la lingüística*, México, UNAM, 1983, p. 34.

Lo anterior nos hace hablar de lengua jurídica, así como se habla de la lengua española, la lengua de los marineros, etcétera.

Con base en las pautas marcadas por Coseriu y Saussure, intentamos definir a la "lengua jurídica" como el sistema de isoglosas o conjunto de oraciones o signos individualizados (habla) comunes en una época y lugar, y que se establecen a partir del consenso, siendo característicos de los estudiosos del derecho.

Esto nos hace ubicar con mayor precisión el problema que nos ocupa, es decir, el del manejo y conocimiento del contenido semántico de los signos dentro del discurso jurídico.

No es suficiente el hecho de pensar que las palabras propias de la lengua o idioma que se emplean comúnmente para ejercer la facultad del habla, y así se dé el fenómeno de la comunicación, tengan o no el mismo significado a nivel jurídico, puesto que aun cuando la lengua jurídica toma prestados de la natural sus signos, palabras, vocablos, éstos no tienen dentro del discurso jurídico el mismo significado; pueden compartir alguno, pero resulta poco probable que compartan todos sus significados a nivel jurídico.

La comunicación llevada a este nivel presenta como "mensaje" a la figura jurídica como tal, o más aún a aquello que el usuario quiere conocer y que obliga a que se realice un estudio lexicográfico.

Un análisis de la norma jurídica precisa ser de contenido lógico-jurídico; sin embargo, para no complicar el proceso de comunicación dentro de nuestro sistema informático, se ha procurado emplear el vocabulario más natural, como natural es precisamente el código que empleamos comúnmente para comunicarnos.

La amplitud, extensión, riqueza, de los signos que conforman la lengua jurídica, reflejan sin duda la riqueza misma del idioma, de la lengua natural, que deviene en técnica-jurídica cuando se le da un nuevo valor o contenido semántico-jurídico a cada significante. Esto es debido a que la materia jurídica como tal es dialéctica, evoluciona conjuntamente con la realidad social, a la que debe ir regulando, por lo que la movilidad la hace prolífica.

La existencia de una nueva figura dentro del universo jurídico provoca la utilización de un nuevo significante, cuyo significado deberá corresponder a aquélla. Así, el legislador, el abogado litigante, el funcionario judicial, el jurista, van armando la propia lengua jurídica de acuerdo con su propia materia de trabajo.

Esto explica el por qué de la existencia de una disciplina o ciencia

cuyo objetivo sea el estudiar los signos jurídicos en su esencia, finalidad y significación dentro del discurso jurídico: la semiótica jurídica.<sup>6</sup>

No es por tanto el trabajo del analista la llave de este mecanismo, si bien su labor permite la integración del sistema; es el "léxico" mismo como instrumento lingüístico el que se encarga de estructurar los signos que empleamos en el discurso jurídico y que facilita el desarrollo del fenómeno de la comunicación.

En resumen, los posibles problemas que entraña la utilización de la lengua jurídica a nivel informático pueden ser reducidos mediante el empleo de este instrumento lingüístico: el léxico.

## 2. *Los problemas de la "lengua" natural*

Por ser la "palabra" la principal categoría gramatical, su incidencia dentro del proceso de comunicación, especialmente cuando éste se representa mediante expresiones matemáticas, hace necesaria la ubicación de las principales características que dicha unidad gramatical puede revestir dentro del discurso y sobre todo en el discurso jurídico.

Decir "características de la palabra" alude a cualidades, mismas que a nivel informático devienen en problemas debido a que la máquina sólo puede reconocer signos y no propiamente significados. Cualidades, en tanto que con pocas palabras puede llegar a expresarse toda una idea; sin embargo, se obliga al interlocutor a realizar una interpretación que le permita conocer el sentido de lo que se le dice con exactitud.

En el área de la informática, este acontecimiento provoca imprecisión, ya que la palabra se torna inexacta e indeterminada, por lo que la economía en el empleo de las palabras hace que la comunicación se vuelva plurívoca.

Es así como el léxico toma su razón de ser como instrumento lingüístico, al tratar de conciliar en nuestro sistema esos problemas que pueden surgir en toda interrogación para convertirlos precisamente en cualidades o ayuda en la comunicación.

Los problemas más comunes que se enfrentan en relación con la lengua son:

a) *Polisemia*, entendida como la asignación de varios significados a un mismo significante o palabra.

b) *Sinonimia*, que es la existencia de varios significantes o palabras

<sup>6</sup> Carrión Wam, Roque, "Semiótica jurídica e informática jurídica", *Revista Latinoamericana de Informática Jurídica*, Venezuela, núm. 1, 1977, p. 53.

que comparten el mismo significado. La sinonimia puede ser absoluta o relativa, es decir, limitada a un cierto contexto.

c) *Antonimia*, consistente en la oposición o contradicción de significados que existe entre diversos significantes.

d) *Analogía*, relación entre palabras o términos que comparten algún núcleo de significado.

## II. INTRODUCCIÓN A LOS SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN DOCUMENTAL

Uno de los problemas que se presenta en todo sistema de recuperación de la información documental es, sin lugar a dudas, el de índole lingüístico, o en otras palabras, aquel problema que se nos presenta para hacer coincidir, a efectos de recuperación pertinente y relacionada, el lenguaje expresado en los documentos almacenados en un banco de datos.

Consideramos este particular como de principal importancia a tratar dado que su existencia incide de manera directa, obstaculizando o interfiriendo en el proceso de comunicación entre el hombre y la máquina que, a resultas, son sujetos principales dentro de un esquema informático documentario.

Ciertamente, "el fin de un sistema automático de información es proporcionar, en relación con una consulta [planteada por el usuario], todas las informaciones pertinentes, y únicamente éstas, contenidas en el sistema";<sup>7</sup> así, podemos decir que este fin lleva *in situ* una doble capacidad: primero, de recuperación, y segundo, de precisión.

Como es sabido, otro de los problemas incidentes en la recuperación de la información en los sistemas automáticos se presenta en virtud de la existencia —en lo que al campo del derecho hace— del carácter especializado y técnico de la lengua jurídica, bien doctrinal, legislativa o jurisprudencial, y su utilización coincidente con los términos de la lengua natural que, digno es de mencionar, da lugar a cierto tipo de ambigüedades lingüísticas que informáticamente se traducirían en el conocido ruido o bien, en su caso, silencio informático.

Precisando el objeto del presente apartado, hemos de mencionar que el problema de coincidencia entre el lenguaje de consulta y el lenguaje documental en la recuperación de la información, actualmente se ha agravado en virtud de la utilización, en los diversos sistemas existentes

<sup>7</sup> Cfr., Taddei Elmi, Giancarlo, *Nociones de informática jurídica documental...*

en la materia, de métodos como el de lenguaje natural —puro o corregido— libre en el plano morfosintáctico, variable —sinonimia— ambiguo —polisemia— en el plano semántico.

Ante esta situación se ha hecho necesario, tanto en la indexación —a palabras clave—, el *full-text* y el análisis —ambos indexación a lenguaje abierto—, el diseñar instrumentos que resuelvan el problema lingüístico.

Las soluciones que se han dado al problema de coincidencia entre el lenguaje documental y el lenguaje de consulta son variadas y van desde los manuales a los automatizados; este particular lo trataremos a continuación.

En principio, hemos de establecer que concebimos un sistema documental como aquel que se integra de los elementos: “consultas, lenguaje o similar, documentos que constituyen en sí el ámbito objetivo, un instrumento para el usuario pero también, en una perspectiva más amplia, podemos conceptualizar tal sistema como ‘una organización de comunicación que incluye al usuario’ mismo.”<sup>8</sup>

Así pues, se puede resolver el problema de la aludida coincidencia, en buena parte, controlando y definiendo previamente el lenguaje de indexación, ya sea tomándolo o no de los documentos fuente. En el caso de un lenguaje cerrado y previamente definido, se reduce o elimina el esfuerzo del usuario en la consulta.

Si se opta por no controlar totalmente el lenguaje, o no hacerlo en lo absoluto, se puede dejar el problema en manos del usuario, quien deberá hacer “suposiciones” sobre el lenguaje documental, tomando en cuenta que puede aprender de sus respuestas y afinar sus consultas.

Sin embargo, existe también la posibilidad de desarrollar ciertos instrumentos a partir del lenguaje documental que incidan sobre sus consultas y muestren al usuario las relaciones semánticas y de otro tipo que se dan entre los conceptos contenidos en este lenguaje, tal como se verá en el caso de otro de los instrumentos lingüísticos llamado *thesaurus*.

Finalmente, la informática tiene a su disposición un arsenal de técnicas y procedimientos propios, que forman parte del mecanismo similar y que, por diversas vías, contribuyen a solucionar el problema de la consulta y la relevancia de las respuestas que obtiene el usuario.

Todo sistema documental hace uso, en realidad, de una mezcla de las opciones básicas anteriormente expuestas, en proporciones y con resultados diversos; de hecho, estas opciones se encuentran estrechamente

<sup>8</sup> Fix Fierro, Héctor, *Informática y documentación jurídica* (tesis recepcional). México, UNAM, 1987.

relacionadas y su punto de referencia se encuentra en el mecanismo similar.

### III. EL LÉXICO EN EL SISTEMA UNAM-JURE

#### 1. *Definición*

Existe una noción del léxico amplia y otra restringida. En sentido amplio, el léxico es el conjunto de signos lingüísticos contenidos en cada uno de los documentos del banco, que en su totalidad conforman una de las partes más esenciales del banco de información: la base documental.

Apoyándose en esta base documental se desprende, desde nuestro punto de vista, la noción en sentido restringido del léxico: instrumento lingüístico de tratamiento del conjunto de signos lingüísticos contenidos en cada uno de los documentos del banco (léxico en sentido amplio), que las organiza, les da cierta estructura, con el fin de resolver los problemas lingüísticos que se le presentan, para que el usuario pueda comunicarse de manera eficaz con la máquina (al recuperar la información pertinente a su interrogación).

Podemos afirmar, entonces, que el léxico en sentido amplio es propio de todo banco de información y que, por el contrario, el léxico en sentido estricto puede o no ser implantado en un banco de información.

El banco de datos legislativos UNAM-JURE ha adoptado al léxico como uno de sus instrumentos lingüísticos de tratamiento u organización de su información documental. Esta información a cuyo tratamiento somete pues el léxico, se encuentra constituida por todos los signos lingüísticos contenidos en los documentos del banco. Esto significa que, de algún modo, tanto las preposiciones como los verbos, así como los adjetivos numerales y palabras extranjeras, por ejemplo, recibirán su tratamiento dentro del léxico.

El léxico va a organizar a las palabras, primero, las más significativas de la lengua (que sean documentalmente relevantes) en unidades semánticas o nociones que constan de diez niveles o subnociones y por medio de las cuales el usuario podrá recuperar la información (por ello se les llaman palabras activas). Este tipo de palabras son los verbos, los sustantivos, los adjetivos y algunos adverbios.

Por otra parte, las preposiciones, conjunciones, artículos y algunos adverbios, cuyo peso semántico es menor (ya que se consideran meros vínculos entre las palabras de una oración, o en su caso meros modifi-



cadores), se tratan en un archivo especial donde, en cierto modo, se les aísla de las palabras activas. De manera que no es posible recuperar la información a través de estas palabras, puesto que se consideran pasivas.

Respecto de los adjetivos numerales, algunos de ellos han recibido un tratamiento en archivos especiales donde se les asigna una subnoción a cada uno: hasta el cinco mil en caso de los cardinales (o arábigos) y hasta el cincuenta en el caso de los romanos. En estos casos dicho tratamiento se estableció desde el origen del banco; por ello, los números ordinales (primero, segundo...), partitivos (onceavo, doceavo...), múltiplos (doble, triple...), que no fueron tratados previamente, se han agrupado en nociones conforme se han presentado en los documentos.

Los adjetivos numerales, a diferencia de las palabras pasivas, sí pueden ser objeto de una interrogación del usuario, en caso de que éste decida emplearlos al comunicarse con la máquina.

## 2. *Objetivo*

El léxico como instrumento lingüístico de tratamiento de la información jurídica documental o lengua jurídica (cuyas características se han señalado), tiene como principal objetivo resolver una serie de problemas lingüísticos para que el usuario, en general, tenga una comunicación satisfactoria con la máquina y, particularmente, para que pueda recuperar toda la información pertinente posible a su interés. El cómo resuelve estos problemas (a través de su organización y su funcionamiento) se intentará explicar en los siguientes apartados.

## 3. *Organización*

La organización o tratamiento de la información documental relevante contenida en los *abstracts* se realiza con base, principalmente, en las palabras activas: aquellos signos lingüísticos a través de los cuales se puede recuperar la información contenida en los documentos. Para activar dichos signos es necesario incorporarlos en una de las nociones, las cuales constituyen la columna vertebral del léxico.

Una noción es una unidad morfológica semántica dividida en subnociones con base, principalmente, en el criterio semántico, relacionado con el criterio gramatical (morfológico y sintáctico).

Entendemos por unidad morfológica semántica el conjunto de palabras que comparten una forma y un significado eje, globalizador, que

permite establecer una cohesión familiar no obstante las mínimas variaciones gramaticales y los cambios semánticos suplementarios que puedan tener las palabras. Así, por ejemplo, la palabra *industria* como unidad morfológica semántica podría estar integrada por las palabras que sean documentalmente pertinentes, de la siguiente manera:

0. Industria industrias	sustantivos
1. Industrial industriales	adjetivos
2. Industrializar industrialización Industrializada industrializadas Industrializado industrializados	formas verbales
3. Industrializadora Industrializadoras	adjetivos (con sufijos)
4. Microindustria microindustrias	sustantivos (con prefijos)

La organización del léxico, basada en esta estructura de las nociones, funciona permitiendo dos tipos de relaciones de las palabras: la automática (relación de equivalencia) y la opcional (relación de extensión). Una explicación más detallada del funcionamiento del léxico se hará en el siguiente apartado.

Los criterios de organización del léxico, es decir, de las nociones, pretenden, pues, resolver los problemas lingüísticos que los datos legislativos contenidos en los *abstracts* nos presentan. ¿Qué clase de criterios son y cuáles son sus excepciones?

#### A. Los criterios de organización

El principal criterio de organización de las palabras de una base documental es el semántico. De acuerdo con la especialización de nuestro banco de información, se concentraría el interés no en la semántica general, sino, sobre todo, en la semántica jurídica. Con ello se pretende integrar una base documental homogénea con fundamento en un conjunto de palabras bien estructuradas semánticamente.

La estructura semántica de las palabras se encuentra organizada, como ya se mencionó, por nociones. Cada noción o unidad morfológica

semántica agrupa clases de palabras de una misma familia semántica, principalmente, y gramatical. Esto es así porque, en general, las palabras con la misma raíz o lexema poseen un significado (unívoco) que permite su agrupación, independientemente de sus accidentes gramaticales: de forma (sean singulares, plurales, femeninos, masculinos, antónimos), o de función (sean verbos, adjetivos, sustantivos o adverbios).<sup>9</sup>

Podríamos afirmar, entonces, que el criterio semántico es la base de formación de las nociones, incluidas —por supuesto— las subnociones (sinónimos, antónimos, polisemia), sólo que éstas estarán divididas, en general, según el criterio gramatical. Por ejemplo, la palabra *informar* como significado eje agrupará en su grupo semántico las siguientes palabras:

0. Informar información	formas verbales
1. Informe informes	sustantivos
2. Informativa informativas	
Informativo informativos	adjetivos
3. Desinformar desinformación	formas verbales

Formalmente, así se agruparían las palabras; sin embargo, la integración de las nociones presentan problemas propios de la lengua que se intentan resolver aprovechando las variaciones semánticas al interior de una misma noción o significado eje. De alguna manera, dichos problemas más son semánticos, pues todo problema lingüístico es, básicamente, una cuestión semántica.

## B. Los problemas lingüísticos

La lengua como fenómeno cultural posee una dinámica que enriquece su desarrollo, es decir, permite que el proceso de comunicación se

<sup>9</sup> En cierto modo, este es el fenómeno conocido como alotaxia: "conjunto de variaciones morfológicas que proceden de las diferentes posibilidades de construcción de las frases al momento del análisis o de la redacción de los documentos: Todos conservando su pertinencia a un mismo campo semántico." Mazed, Guy y Claude Bernad, *Traitement par ordinateur des variables allotaxiques*, Institut de Recherches et d'Etudes pour le Traitement de l'Information Juridique de l'Université de Montpellier, France, 1976, p. 12.

esté continuamente recreando. La lengua, en este sentido, se mantiene no sólo vigente sino cambiante.

Por su parte, el léxico, como instrumento de organización de la lengua, se enfrenta a dos fenómenos: el primero rígido, en tanto que va a reagrupar las palabras con base en las características de la lengua misma, es decir, se trata de un sistema de reglas que el léxico debe respetar. El segundo fenómeno es el dinámico, ya que la lengua, por ser un hecho social —como ha quedado indicado—, se encuentra en constante evolución, lo cual obliga al léxico a actualizarse periódicamente.

El léxico, pues, tratará de agrupar a las palabras tomando en cuenta esta dinámica de la lengua, y en particular de la “lengua” jurídica, cuya semántica —nos atreveríamos a decir— es más rígida por utilizar una terminología muy precisa. Esta característica del derecho determinará que el principal criterio de agrupamiento de las palabras (como ya se ha mencionado) sea el jurídico.

En un apartado anterior<sup>10</sup> se han mencionado otras características de la “lengua” jurídica. Dichas características presentan una serie de problemas en la agrupación de las palabras que el léxico, como instrumento lingüístico de organización, tendrá que resolver con base en los criterios señalados, para que esta especial comunicación entre usuario y máquina se lleve a cabo de la manera más eficiente posible.

Al adoptar el “lenguaje abierto” para la redacción de los *abstracts*, el análisis tiene cierta libertad para escribir a su modo las ideas contenidas en el documento fuente, siempre y cuando no las modifique. Por consiguiente, el usuario, al intentar comunicarse con la máquina, lo podrá hacer con “lenguaje abierto”, es decir, escogiendo las palabras que considere a su interés documental.

Esta “libertad” del analista para redactar sus *abstracts* se ve limitada por los conceptos tan precisos que se contienen en la legislación y por razones técnicas ya determinadas de estructura de los párrafos y frases de los *abstracts*.<sup>11</sup> Ante esta situación, el analista deberá tener cuidado en la manera de redactar las ideas contenidas en los documentos fuente, ya que sin estar sujeto a emplear las mismas palabras del documento, debe redactar su *abstract* de modo que los conceptos puedan ser recuperados por el usuario. De esta manera, el analista encontrará palabras con una carga semántica jurídica preponderante que no podrá sustituir. Sin embargo, en la explicación misma de los conceptos (ex-

<sup>10</sup> Ver apartado I.1 de este artículo.

<sup>11</sup> Ver “La metodología del *abstract* legislativo”, trabajo presentado por Enrique Cáceres Nieto en esta misma publicación.

plícitos e implícitos) sí podrá utilizar otras palabras siempre que no modifique sustancialmente el concepto. Con base en ello, pues, podrá cambiar el tiempo del verbo, utilizar sinónimos, polisemias, antónimos o análogos.

Lo anterior obliga al sistema UNAM-JURE emplear instrumentos lingüísticos (léxico y *thesaurus*) que sirvan de control de las palabras que ingresen al banco, para que, independientemente de las empleadas tanto por el analista como por el usuario, los conceptos contenidos en los *abstracts* se puedan recuperar exhaustiva y pertinentemente. Para ello, el léxico intentará controlar, principalmente, los cambios verbales, los sinónimos y algunos casos de polisemia y de antonimia.

#### a. *Los cambios verbales*

El analista, al redactar la idea de "creación de una empresa", podrá modificar el tiempo del verbo: "se creó una empresa", "se ha creado una empresa", "se creará una empresa". Esta cuestión el léxico la resuelve agrupando en una misma subnoción todas las formas verbales de la unidad morfológica semántica correspondiente. Por ejemplo:

0. Creación creaciones creo crearé crearán creado.

De tal modo que el usuario podrá plantear su interpretación por cualquiera de las formas verbales y recuperar el documento, ya que una de las funciones del léxico es relacionar de manera automática todas las palabras agrupadas en una misma subnoción.

#### b. *Los sinónimos*

Cuando se analiza un reglamento de cementerios, podrá el analista redactar sus *abstracts* poniendo: Reglamento de panteones. El léxico ha solucionado este problema agrupando en la misma subnoción a las palabras con el mismo significado. De manera que si el usuario pregunta por cualquiera de las dos palabras se recupere el mismo documento, ya que los sinónimos —al igual que las formas verbales— se integran, pues, en una misma subnoción:

0. Panteón panteones cementerio cementerios.

Nótese que en este caso se rompe con el criterio general de agrupar en una misma subnoción las palabras con la misma raíz o lexema. Sin embargo, por ser sinónimos perfectos se tratan como equivalentes.

A diferencia, los análogos tienen que separarse en nociones distintas, a pesar de compartir un significado similar, como sería el caso de las palabras: automóvil, vehículo, carro, coche, cuya recuperación sólo sería posible a través del *thesaurus*, que permite relacionar palabras de diferentes nociones.

### c. *La polisemia*

Son polisémicas aquellas palabras que tienen varios significados. El analista podrá utilizar éstos en su redacción: en un reglamento de reclusorios podrá desarrollar el conjunto de información correspondiente a las obligaciones de las "prisioneras", pudiendo utilizar en su lugar la voz "presas", ya que estaría empleando una palabra con varios significados: persona encarcelada, animal atrapado, y muro que impide o desvía el curso de un río, arroyo o canal.

En el caso planteado, se podría establecer su significado preponderante por el contexto; sin embargo, al momento de agrupar la palabra en nuestro léxico se carece del contexto en el cual la palabra fue utilizada, por lo cual sólo se le identifica a la palabra como polisémica (ruptura semántica parcial), y como la palabra "presa" comparte un significado común con las demás palabras, entonces se agrupa en esa noción, sólo que ocupando una subnoción de la misma:

#### 0. Prisión prisiones

1. Prisionero prisioneros prisionera prisioneras

2. Preso presos

3. Presa presas

Al agrupar a las palabras en las nociones, debemos tener en cuenta la naturaleza especializada de la información; en este sentido podríamos afirmar que cada una de las palabras tienen un contexto jurídico implícito. Por ello, en el caso de "presas" se optó por agruparla en la unidad morfológica semántica que corresponde a "prisiones" para no perder el sentido de "prisioneras" al momento que el usuario plantee una pregunta general. Y, por otra parte, aislada de "presos" para que en una pregunta concreta se recuperen los otros dos sentidos que, en este caso, vendrían a ser un tanto suplementarios al significado jurídico común de la noción.

Podemos adelantar respecto del funcionamiento del léxico, que éste permite relaciones de las palabras estructurales en toda la noción (para

ampliar la búsqueda de la información) y relaciones de sólo una subnoción (para un tipo de información muy concreto).

Otro caso que nuestro léxico trata, en relación con las palabras polisémicas, se presenta cuando el analista al redactar su ficha sobre la creación, por ejemplo, de una empresa de participación estatal, escribe "Constitución de la empresa...". Empleando así la palabra "constitución" en el contexto tendrá que agruparla como palabra polisémica. Para ello, cabe advertir que la palabra "constitución" podría agruparse en las nociones de "constitución" o en la "constitucional". Técnicamente no se podría incorporar en ambas, ni informáticamente conveniría ubicarla en una de ellas dejando incompleta la otra. Lo que se adoptó fue hacer de este tipo de palabras una noción puente entre ambas, advirtiendo al usuario que en cada una de las nociones existe el aislamiento semántico y, por tanto, que la palabra "constitución" ha sido aislada.

#### 0. Constituir 0. Constitución 0. Constitucional

Sin embargo, es necesario mencionar que al escribir este artículo ya se estaba trabajando en resolver el problema de incluir una subnoción en dos o más nociones.

Con base en el contexto jurídico implícito y su comprobación a través de una investigación de las fichas de análisis, el criterio anterior puede ser modificado. La palabra "importar", por ejemplo, que podría agruparse en las nociones de "importación" o de "importante", se decidió agruparla en la primera, puesto que en una muestra de 1123 palabras en más del 95% tenía en su contexto el significado de importación.

Por último, tenemos el caso en que todas las palabras de una noción comparten más de un significado idéntico (polisemia paralela). El tratamiento que se le da en léxico es agrupando en una misma noción a todas las palabras:

- 0. Circulación circulaciones
- 1. Circular circulares circulando
- 2. Círculo círculos.

#### d. *La antonimia*

La antonimia es un problema lingüístico que el léxico ha tratado de resolver uniendo y separando las palabras según sea la clase de antonimia: morfológica o general.

Cuando es una *antonimia morfológica o derivacional* se unirán o agruparán a las palabras en una subnoción de la familia o unidad semántica correspondiente. Esto es así porque desde el punto de vista de la recuperación documental automática, los antónimos expresan dos aspectos de una misma realidad, sin que por ello la semántica sea afectada.<sup>12</sup> Por ejemplo, se puede variar la forma sin alterar el sentido cuando el analista hace cambios sintácticos:

No hay disciplina = hay disciplina.

Y en la noción correspondiente quedará como sigue:

- 0. Disciplina disciplinas
- 1. Disciplinarse disciplinar
- 2. Interdisciplinario interdisciplinaria  
Interdisciplinarios interdisciplinarias
- 3. Multidisciplinario multidisciplinaria
- 4. Indisciplina.

Cuando es una *antonimia general* se separarán las palabras en nociones distintas, puesto que no pertenecen a la misma familia morfológica y no comparten significados similares. Por ejemplo:

Hombre/mujer

#### 4. *Funcionamiento*

Las diversas nociones que conforman estructuralmente el léxico constituyen la base de su funcionamiento. Tomaremos la siguiente noción para explicar cómo funciona el léxico.

- 0. Industria industrias
- 1. Industrial industriales
- 2. Industrializar industrialización industrializada  
industrializadas industrializado industrializados
- 3. Industrializadora industrializadoras
- 4. Microindustria microindustrias.

<sup>12</sup> Ver Mazed, Guy y Claude Bernard, *op. cit.*, p. 26.



El léxico funciona permitiendo la relación de las palabras al momento que el usuario plantea una interrogación. Existen dos tipos de relación de las palabras que el léxico permite: la automática (relación de equivalencia) y la opcional (relación de extensión).

#### A. Relación de equivalencia

Esta relación conjunta las palabras que se encuentran agrupadas en una subnoción, realizándolo de manera automática. Así, si el usuario planteara la pregunta sobre la *industria*, la búsqueda abarcará también las *industrias*. De esta manera se recuperarán todos los *abstracts* donde se encuentren las palabras *industria* e *industrias*, aunque ésta no haya sido puesta al plantear la interrogación.

Este funcionamiento permite al usuario una recuperación de la información de manera directa y, principalmente, sobre materias muy particulares.

#### B. Relación de extensión

Esta relación conjunta las palabras que se encuentran agrupadas en toda la noción, pero opcionalmente, es decir, sólo si el usuario decide ponerlo en funcionamiento.

Si de la noción que citamos le interesa al usuario no sólo saber lo que exista en el banco de datos sobre la *industria*, que sería muy particular, sino todo lo que haya sobre el fenómeno industrial, tendría que utilizar en su interrogación el comando & después de cualquier palabra de la noción correspondiente (*industrial&*, por ejemplo) para que la máquina relacione todas las palabras de la familia. Ello permitirá al usuario recuperar documentos donde haya no sólo la palabra *industrial* e *industriales* (ya que esta relación de palabras es automática por estar en la misma subnoción), sino también documentos donde se contienen las palabras de las otras subnociones: *industria(s)*; *industrializar*, *industrialización*, *industrializada(s)*, *industrializado(s)*; *industrializadora(s)*; *microindustria(s)*.

El tipo de relación automática, entonces, permite al usuario realizar interrogaciones sobre materias muy concretas. Teniendo, además, la opción de ampliar su pregunta al hacer uso del segundo tipo de relación (sin que necesariamente tenga que seguir este orden).

Con este funcionamiento del léxico, los problemas lingüísticos, previamente identificados, y que se intenta resolver son: la sinonimia, la polisemia (salvo la aislada) y la antonimia derivacional o morfológica.

Los casos que el *thesaurus* trata son la polisemia aislada y la antonimia general.

### 5. *Tratamientos especiales*

En este apartado se intenta explicar los casos que no son tratados con base en los criterios generales establecidos, sino que por la naturaleza semántica gramatical de las palabras (conjunciones, preposiciones, artículos, pronombres y algunos adverbios), o bien por ser palabras sin parentesco ni unidad morfológica (sustantivos propios, adjetivos numerales, siglas, latinismos, anglicismos), el léxico les da un tratamiento especial.

#### A. Por la naturaleza semántica gramatical de las palabras

Las conjunciones, preposiciones, artículos, pronombres y algunos adverbios son palabras con poco peso semántico dentro de una oración, ya que sólo tienen la función de ser un vínculo, un nexo, entre las demás palabras de una frase, como sería el caso de las conjunciones; o bien, tienen las funciones de introducir complementos como las preposiciones; de ser un tipo especial de adjetivo como los artículos, y de ser pronombres según el contexto.

Estas palabras forman parte del léxico en sentido general, pero el léxico en sentido estricto (como instrumento lingüístico) les da un tratamiento especial por tener poco peso semántico y casi nula relevancia documental. Lo que nuestro léxico hace, pues, es agruparlas en un archivo especial, aislado del resto de las nociones, lo cual impide que el usuario pueda recuperar los documentos a través de su empleo. De modo que todas las palabras incorporadas en este archivo especial sean consideradas pasivas, informáticamente hablando.

#### B. Por ser palabras que carecen de parentesco morfológico

Las palabras carecen de un parentesco morfológico, en el contexto de este trabajo, por ser palabras que se caracterizan por tener cierta autonomía morfológica que impide su agrupamiento estricto por familias. Sin embargo, son integradas en nociones a pesar, pues, de no te-

ner ninguna relación morfológica ni, en la mayoría de los casos, semántica. El único criterio, en todo caso, que permite su agrupamiento es el gramatical, y su nivel de recuperación sería a través de la subnoción (relación de equivalencia) por ser información muy particular.

a) Los sustantivos propios, sean de pueblos o de personas (aunque suelen intercambiarse), por lo general están compuestos de varias palabras y con significado único. De manera que se optó por unirlos con guiones para que la máquina las reconociera como una sola palabra y así conservar su significado.

Nuestro léxico agrupa, entonces, en cada subnoción un sustantivo propio, y en el caso de los pueblos se les agrega sus gentilicios:

0. Nayarit nayarita nayaritas
1. Miguel-Hidalgo-y-Costilla
2. Perú peruana peruano peruanas peruanos
3. Ojo-de-liebre
4. Pátzcuaro
5. Jalisco jalisciense jaliscienses
6. Taxco Taxco-de-Aragón
7. Brasil brasileño brasileña brasileños brasileñas
8. Rosamorada
9. Santa-Teresa
10. Emiliano-Zapata

El usuario, en este caso, podrá recuperar este tipo de información (por ser muy concreta) utilizando la relación de equivalencia. Aunque podría el usuario optar por la relación de extensión, es decir, de toda la noción, sin embargo, el riesgo de ruido sería muy alto.

b) Los adjetivos numerales llamados cardinales y romanos han sido organizados previamente por el léxico (desde la creación de nuestro banco de datos) en archivos especiales donde a cada número le corresponde una subnoción: en el archivo de los cardinales hasta el cinco mil y de los romanos hasta el cincuenta ("L"). Esto permite al usuario plantear preguntas muy concretas: reformas a artículos y fracciones, por ejemplo.

Los adjetivos numerales llamados ordinales partitivos y múltiples se agrupan, al igual que los demás adjetivos numerales, uno por cada subnoción; con la diferencia que éstos no tuvieron un tratamiento previo en archivos especiales, sino que conforme han aparecido en los documentos se han creado las nociones correspondientes.

c) Las siglas se agrupan en nociones correspondientes, una sigla por subnoción. Esto permite al usuario tener acceso a los documentos no sólo a través del nombre de la disposición o institución sino a través de sus siglas. Por ejemplo: Ley del Impuesto al Valor Agregado podrá utilizar Ley IVA, o en lugar de Instituto Mexicano del Seguro Social podrá plantear su pregunta por IMSS.

Hay casos en que una sigla puede tener, como palabra, otro significado, por ejemplo: CREA puede ser una sigla (Consejo Nacional de Recursos para la Atención de la Juventud) o bien una forma verbal (crear o creer). En este caso, por el contexto jurídico implícito, se optó por conservar la palabra como sigla, por la importancia del organismo al que representa.

d) Las palabras extranjeras que a veces utiliza el legislador o que son empleadas por el analista en caso de conceptos implícitos, son agrupadas en nociones donde cada palabra extranjera tenga un significado que corresponda a una unidad semántica ya establecida. Por ejemplo: el latinismo *ad-valorem* se ha integrado a la noción de *valor*.

#### IV. CONCLUSIONES

El léxico como conjunto de palabras contenidas en un banco de datos corresponde a una noción amplia, y el léxico como instrumento de organización de las palabras por su significado, principalmente, se refiere a una noción estricta. Ambas nociones se interrelacionan, pero en algunos casos se les confunde.

El léxico como instrumento lingüístico tiene por objeto servir de intermediario entre el usuario y la máquina, para que aquél pueda comunicarse con ésta, recuperando toda la información pertinente a su interrogación.

Para que el usuario pueda tener una comunicación eficaz con la máquina, el léxico se encarga de resolver una serie de problemas lingüísticos propios de la lengua, organizando a las palabras en unidades morfológicas semánticas llamadas nociones, las cuales se caracterizan por tener diversas subnociones con cierta independencia semántica.

El principal criterio de organización o de agrupamiento de las palabras es el semántico. Y por tratarse de un banco de datos especializado nuestra herramienta principal es la semántica jurídica.

Tanto el analista como el usuario pueden utilizar cualquier palabra en el momento de hacer su análisis o plantear su interrogación. Esto nos permite hacer un paralelismo entre la manera como el analista

estructura su información y el modo como el usuario puede recuperar la información.

El analista estructura su información a dos niveles: párrafo y frase. El usuario puede plantear, también, su interrogación a dos niveles (puesto que el funcionamiento del léxico se lo permite): general (a nivel de toda la noción) y particular (a nivel de una subnoción).

El nivel párrafo viene a ser para el analista lo que el nivel general es para el usuario; sólo que para el primero la generalidad que se manifiesta a nivel párrafo se refiere a una unidad de información, y para el segundo se refiere a una unidad morfológica semántica, es decir, a una noción del léxico.

Una de las posibilidades que tiene el léxico de funcionar más ampliamente es extendiendo su capacidad de relacionar las palabras, no sólo a las que se encuentran a nivel subnoción o a nivel de toda la noción, sino las que se encuentran en distintas nociones tanto a nivel subnoción como a nivel de la noción completa, incluso discriminando subnociones, permitiendo al usuario formar cadenas o arborescencias semánticas pertinentes a su interés documental.

El desarrollo del léxico en el sistema UNAM-JURE en estos casi siete años de existencia nos permite afirmar que, no obstante los cambios de criterios habidos en la agrupación de las palabras, éstos no han sido de esencia y su funcionamiento tiene ya una etapa de consolidación que en los próximos años, esperamos, se verá confirmada para bien del desenvolvimiento de la informática jurídica documental en México y en América Latina.